# Does Normalized Whole Brain Volume in Dementia Patients Across Groups Change Over Time?

## Introduction

According to Sophia Z. Liu et al. a decrease in ***Normalized Whole Brain Volume*** (**NWBV**) is associated with dementia in both men and women.[6] Liu et al. note that **NWBV** generally decreases over time and with age. If we can show that **NWBV** decreases across visits for patients with dementia in a more pronounced fashion than it does for patients without then it would be beneficial in detecting dementia through brain volume scans. We would need a baseline scan for comparison. However, this could be instituted by ensuring all patients get a baseline scan after a certain age.

## Research Question

1. Does the rate of ***Normalized Whole Brain Volume*** (**NWBV**) change over time across demented and non-demented patients and what is the impact of a patient's dementia status on the change in **NWBV**?

## Data Cleaning and Data Wrangling

The dataset is generally clean and requires few modifications. The dataset consists of 15 **columns** and 294 **rows** in a comma-separated values file. The data has 15 missing values in the **SES** column and 1 missing value in the **MMSE** column. We have opted to drop the records with missing entries. The **SUBJECT_ID** column is a unique identifier for each record. We made some minor changes by renaming the columns for the sake of stylistic consistency and to ensure simpler code. Additionally, we converted all columns with continuous variables to the number format.

***Table 1*: Data Columns and Descriptions**

| Column Name in Original Data | Column Name in Data Frame | Description |
|---|---|---|
| Subject ID | SUBJECT_ID | Unique patient identifier. |
| MRI ID | MRI_ID | Unique MRI exam identifier. |
| Group | GROUP | Categorical variable indicating whether a patient is "Demented, Non-Demented, or Converted." |
| Visit | VISIT | Categorical variable indicating the doctor's visit order. |
| MR Delay | MR_DELAY | Continuous variable for the number of days between two medical visits. |
| M/F | SEX | Categorical variable for the sex of the patient. |
| Hand | HAND | Categorical variable for the dominant hand of the patient. |
| Age | AGE | Continuous variable indicating the age of the patient at the time of the MRI. |
| EDUC | EDUC | Continuous variable indicating years of education a patient has received. |
| SES | SES | Categorical variable for socio-economic status. Higher is "better." 1: less than high school, 2: high school grad, 3: some college, 4: college grad, 5: beyond college. |
| MMSE | MMSE | Mini-Mental State Examination |
| CDR | CDR | Categorial variable for "Clinical Dementia Rating" (0=nondemented, 0.5 – very mild dementia, 1 = mild dementia, 2 = moderate dementia) |
| eTIV | ETIV | Continuous variable indicating Estimated Total Intracranial Volume. |
| nWBV | NWBV | Continuous variable indicating "Normalized Whole Brain Volume." This measure accounts for variation in individual head size. |
| ASF | ASF | Continuous variable for the Atlas Scaling Factor. This factor is used to correct for regional head-size variation. |

## Descriptive Statistics and Exploratory Data Analysis

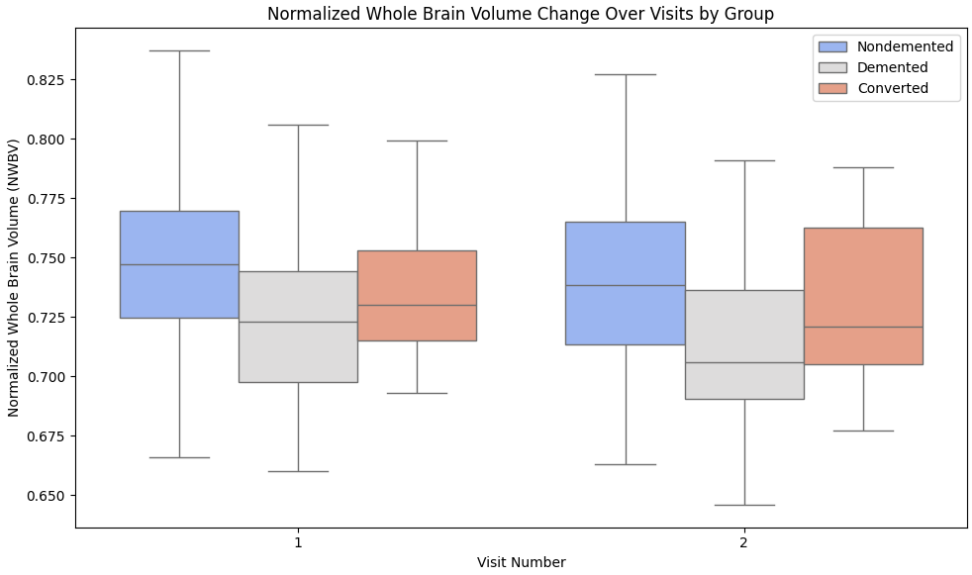
***Table 2*: Descriptive Statistics of Continuous Variables (Post-Cleaning)**

| Column | Count | Mean | Std | Min | 25% | 50% | 75% | Max |
|---|---|---|---|---|---|---|---|---|
| VISIT | 279 | 1.491039 | 0.500818 | 1 | 1 | 1 | 2 | 2 |
| MR_DELAY | 279 | 353.200717 | 404.525563 | 0 | 0 | 0 | 673 | 1707 |
| AGE | 279 | 76.394265 | 7.775314 | 60 | 71 | 76 | 82 | 98 |
| EDUC | 279 | 14.684588 | 2.901404 | 6 | 12 | 15 | 16.5 | 23 |
| SES | 279 | 2.491039 | 1.128008 | 1 | 2 | 2 | 3 | 5 |
| MMSE | 279 | 27.318996 | 3.435222 | 15 | 26 | 29 | 30 | 30 |
| CDR | 279 | 0.284946 | 0.382141 | 0 | 0 | 0 | 0.5 | 2 |
| ETIV | 279 | 1479.433692 | 176.748150 | 1106 | 1348.5 | 1461 | 1568.5 | 2004 |
| NWBV | 279 | 0.731703 | 0.038061 | 0.646 | 0.702 | 0.732 | 0.757 | 0.837 |
| ASF | 279 | 10.67122 | 0.138758 | 0.876 | 1.1185 | 1.202 | 1.3015 | 1.587 |

We can see from ***Table 2*** above that the data is predominantly made up of older adults with an average age of approximately **76** years old. Furthermore, we can tell that the respondents are well educated with an average of **14** years

of education and a minimum of **6** years. Most of the respondents have had approximately **1** visit or more, this is worth noting because without this we would be unable to do a longitudinal analysis. The respondents are relatively diverse in terms of socioeconomic status as indicated by the mean **SES** score of **2.49**. The mean **CDR** and the standard deviation for **CDR** reveal that the dataset contains patients in varying stages of the dementia diagnosis. Let's explore our data and research question further by drawing a boxplot.

***FIGURE 1*: BOXPLOT OF CHANGE IN NORMALIZED WHOLE BRAIN VOLUME OVER VISITS BY GROUP**




We can see from ***FIGURE 1*** above that the dementia status of a participant may impact the participants ***Normalized Whole Brain Volume*** (**NWBV**). The "non-demented" patients have a higher mean **NWBV** regardless of **VISIT** number. The mean **NWBV** for "demented" patients drops drastically across visits. Let's look at these results in a tabular format below.

***TABLE 3*: IQR, MEAN, & MEDIAN FOR NORMALIZED WHOLE BRAIN VOLUME OF PATIENTS BETWEEN VISITS ACROSS GROUPS**

| GROUP | VISIT | IQR | MEAN | MEDIAN |
|---|---|---|---|---|
| Converted | 1 | 0.03775 | 0.737786 | 0.7300 |
| | 2 | 0.05775 | 0.728417 | 0.7210 |
| Demented | 1 | 0.04650 | 0.723482 | 0.7230 |
| | 2 | 0.04600 | 0.712764 | 0.7060 |
| Nondemented | 1 | 0.04500 | 0.746125 | 0.7470 |
| | 2 | 0.05150 | 0.737671 | 0.7385 |

We can see from ***TABLE 3*** that the **NWBV** of all patients across all groups decreases with time. This may reveal the normal ravages of time on brain volume. However, it is worth noting the means of the **NWBV** for converted and demented patients is especially pronounced. Based on this let's formulate a hypothesis and run some tests to see if our data satisfies the assumptions for a mixed-effect **ANOVA**.

## HYPOTHESES

***TABLE 4*: HYPOTHESES (USING A SIGNIFICANCE VALUE OF 0.05)**

| HYPOTHESES FOR RESEARCH QUESTION | | |
|---|---|---|
| $H_{null}$ | : | There is no difference in the rate of change of NWBV over time across patients diagnosed as demented and those diagnosed as non-demented. In other words, dementia has no impact on the change in brain volume. |
| $H_{alternative}$ | : | The rate of change of NWBV over time across patients diagnosed as demented and those diagnosed as non-demented differs and is statistically significant. In other words, dementia impacts the change in brain volume over time. |

## Checking Assumptions for Two-Way Mixed-Effect Anova

Before we define and fit a mixed-effects model let's ensure that the assumptions for such a model are satisfied. Let's start of by testing for normality as this is one of the assumptions for a mixed-effect **Anova**. As we can see from the table below all the subgroups satisfy the normality assumption as the p-values for each subgroup is greater than 0.05.

***Table 5*: Normality Test for Each Subgroup**

| Group | Visit | P-Value | Exact P-Value | Sub-Group is Normal (True / False) |
|---|---|---|---|---|
| Converted | 1 | **p** > 0.05 | 0.3414671116115821 | True |
| | 2 | **p** > 0.05 | 0.5023515660175943 | True |
| Demented | 1 | **p** > 0.05 | 0.8777173909634739 | True |
| | 2 | **p** > 0.05 | 0.2948692647530252 | True |
| Nondemented | 1 | **p** > 0.05 | 0.771978318043403 | True |
| | 2 | **p** > 0.05 | 0.5108552369529133 | True |

Furthermore, our **test statistic for Levene is 0.296** and we have a **p-value > 0.05**. These results reveal that there are no significant differences in variances of **NWBV** across groups (converted, demented, and nondemented). This satisfies the homogeneity of variances assumption for our **Anova**. The **homogeneity of covariances** condition is also satisfied as our sphericity test (***Mauchly***) has a **p-value = 1**. In the next section we will run the **Anova** using the following parameters: the ***dependent variable*** will be **NWBV**, the ***within-subject factor*** will be **VISIT**, the ***between-subject factor*** will be **GROUP**.

## Results of the Two-Way Mixed-Effects Anova

***Table 6*: Two-Way Mixed-Effects Anova (DV = NWBV, within=VISIT, between=GROUP)**

| Source | SS | DF1 (GROUP) | DF2 (ERROR) | MS | F | p-unc (unadjusted) | $n_p^2$ | eps |
|---|---|---|---|---|---|---|---|---|
| **GROUP** | 0.033179 | 2 | 134 | 0.01659 | 6.384264 | 0.002247 | 0.086998 | |
| **VISIT** | 0.006225 | 1 | 134 | 0.006225 | 89.37623 | $1.45 \times 10^{-16}$ | 0.400115 | 1 |
| **Interaction** | 0.000227 | 2 | 134 | 0.000114 | 1.630389 | 0.199714 | 0.023756 | |

The results of our **Anova** shown in ***Table 6*** above reveal that there are significant differences in **NWBV** across the three different groups. We can see that the effect size shown in the $n_p^2$ column (partial Eta squared aka $\eta_p^2$ value) for the **GROUP** factor is responsible for approximately 8.7% of variance in the **NWBV**.[2] This is a **moderate effect size**. Furthermore, the **p-value < 0.05** and this indicates that the difference in **NWBV** across the three groups is statistically significant. Additionally, we can see from the partial Eta squared value for the **VISIT** factor that this factor has a **large effect size**, it *explains approximately 40% of the variance in NWBV*. The **p-value** for this factor is also ***less than 0.05***. This provides evidence to support the change in **NWBV** from **VISIT1** to **VISIT2** irrespective of the **GROUP** of the patient. We have already demonstrated in our **EDA** that the **NWBV** decreases between visits across **GROUPS** over visits (**VISIT)**. The results of the **Anova** also reveal the **interaction effect** between **GROUP** and **VISIT**. We can see from the **p-value** for this which is ***greater than 0.05*** that the interaction between **VISIT** and **GROUP** is not statistically significant. In other words, **NWBV** does not change over visits (**VISIT**) across **GROUP** (demented, nondemented, and converted). Finally, we can see from the **effect size** as shown by the partial Eta square value ($n_p^2$ column) that the interaction effect only explains a little over 2% of the variance in **NWBV**. Next, we will run **pairwise T-tests** to determine how specific groups (demented, nondemented, or converted) differ with respect to their mean **NWBV**. The results of our **post-hoc pairwise T-test**, shown in ***Table 7***, below reveal that the uncorrected p-value between VISIT1 and VISIT2 is less than **0.05**, indicating a strong change in **NWBV** over visits. The effect size is approximately 0.251 indicating a **medium effect**. We also know from the post-hoc test that the change in **NWBV** between "demented" and non-demented" patients is statistically significant as the p-value is ***less than 0.05***. Furthermore, we know that the effect size is -0.637. This indicates that the NWBV is lower in "demented" patients. The $\mathbf{BF_{10}}$ column indicates the **bayes factor**, the higher the values here the stronger the evidence against our null hypothesis. I have highlighted the relevant lines in the table below. We can clearly see that the **NWBV** between "demented" and "non-demented" patients over visits (**VISIT**) decreases significantly. We can also see as shown by the **bayes factor** value of 99.737 in the last row that the decrease is **NWBV** is greater in **VISIT2** than in **VISIT1**.

**TABLE 7: POST-HOC PAIRWISE T-TESTS**

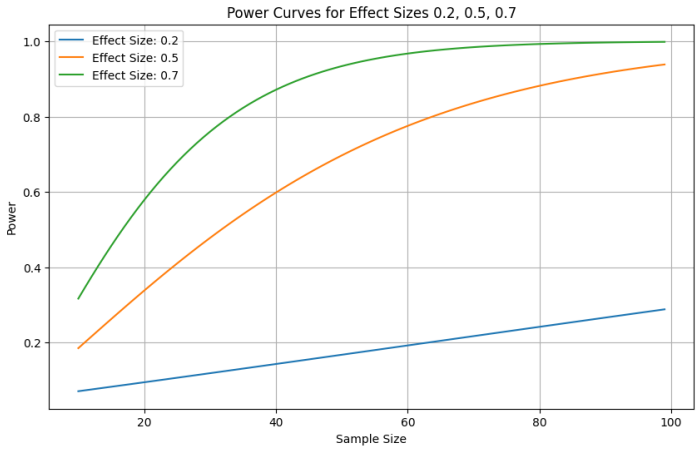
| CONTRAST | VISIT | A | B | PAIRED | PARAMETRIC | T | DOF | ALTERNATIVE | P-UNC | $BF_{10}$ | HEDGES |
|---|---|---|---|---|---|---|---|---|---|---|---|
| VISIT | - | 1 | 2 | TRUE | TRUE | 9.410382 | 136 | two-sided | 1.67E-16 | $3.512 \times 10^{13}$ | 0.250755 |
| GROUP | - | Converted | Demented | FALSE | TRUE | 1.43556 | 15.52273 | two-sided | 0.170972 | 0.698 | 0.471005 |
| GROUP | - | Converted | Nondemented | FALSE | TRUE | -0.64752 | 15.83231 | two-sided | 0.526572 | 0.361 | -0.18844 |
| GROUP | - | Demented | Nondemented | FALSE | TRUE | -3.61969 | 121.9186 | two-sided | 0.000431 | 59.81 | -0.63674 |
| VISIT * GROUP | 1 | Converted | Demented | FALSE | TRUE | 1.467275 | 15.43273 | two-sided | 0.162384 | 0.724 | 0.484403 |
| VISIT * GROUP | 1 | Converted | Nondemented | FALSE | TRUE | -0.46502 | 15.99496 | two-sided | 0.648188 | 0.333 | -0.13384 |
| VISIT * GROUP | 1 | Demented | Nondemented | FALSE | TRUE | -3.34307 | 122.4173 | two-sided | 0.0011 | 26.387 | -0.58634 |
| VISIT * GROUP | 2 | Converted | Demented | FALSE | TRUE | 1.367894 | 15.78933 | two-sided | 0.1905 | 0.648 | 0.440954 |
| VISIT * GROUP | 2 | Converted | Nondemented | FALSE | TRUE | -0.81148 | 15.61469 | two-sided | 0.429286 | 0.397 | -0.23983 |
| VISIT * GROUP | 2 | Demented | Nondemented | FALSE | TRUE | -3.78373 | 120.5262 | two-sided | 0.000242 | 99.737 | -0.66947 |

## ANALYSIS

Our results can be summarized as follows. The NWBV between "demented" and "non-demented" patients decreases significantly over visits. In other words, there is a greater decrease in NWBV for demented patients over visits when compared to non-demented patients. There is also a strong interaction effect between a patient's dementia diagnosis and their visit. The NWBV for patients diagnosed with dementia decreases more drastically over visits than it does for patients without dementia. We ran a further test which utilized a two-way mixed-effects **ANOVA** using the Mini-Mental State Examination (MMSE) as the dependent variable. The results of this **ANOVA** can be found *in the Jupyter Notebook and have not been documented here due to a 4-page limit*. In brief, these results confirmed that there are statistically significant differences in MMSE between "demented" and "non-demented" patients. Approximately, 46% of the variance in MMSE arises from a patient's dementia status. Furthermore, there is an interaction effect between a patient's dementia diagnosis and their MMSE score. This interaction is statistically significant and explains 5% of the variance in MMSE scores. Finally, the post-hoc pairwise T-test revealed that MMSE scores for patients with dementia over visits are statistically different from those for patients without dementia. The MMSE scores are much lower for patients with dementia than those without. ***See section 4.5 in the Jupyter notebook*** for the relevant **ANOVA** and **POST-HOST PAIRWISE T-TEST** tables.

## POWER ANALYSIS

**FIGURE 2: POWER CURVES FOR EFFECT SIZES 0.2, 0.5 AND 0.7**




$alpha = 0.05$
$power = 0.91$
$effect\ size = 0.7$
$cumulative\ alpha\ probability = 1 - (alpha/2)$
$z\ alpha_2\ value = 1.959963984540054$
$cumulative\ beta\ probability = 1 - power = 1 - 0.91 = 0.09$
$z\ beta\ value = 1.3407550336902165$

$$sample\ size\ pe\ group = 2 \times \left(\frac{z\ alpha_2\ value + z\ beta\ value}{effect\ size}\right)^2$$

$$= 2 \times \left(\frac{1.959963984540054 + 1.3407550336902165}{0.7}\right)^2$$

$$= 2 \times \left(\frac{3.30071901823026}{0.7}\right)^2 \approx 2 \times 22.23 = 44.46 \approx 44$$

$\therefore$ *for a theoretical experiment with* $alpha = 0.05, power = 0.91,$ *and* $effect\ size = 0.7$ *we need 44 particiapnts in each group.*

The theoretical experiment with a ***power = 0.91***, ***alpha = 0.05***, and ***effect size = 0.7*** would need to have a sample size of approximately 44 participants per group. ***Calculations shown above and in Jupyter notebook section 5***.

## REFERENCES


1. Buckner, Randy L., Denise Head, Jamie Parker, Anthony F. Fotenos, Daniel Marcus, John C. Morris, and Abraham Z. Snyder. "A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume." *Neuroimage* 23, no. 2 (2004): 724-738.

2. Daines, Ramon. n.d. 'LibGuides: Statistics Resources: Partial Eta Squared'. Accessed 4 April 2024. https://resources.nu.edu/statsresources/eta.

3. Marcus, Daniel S, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner. 2007. "Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults." *Journal of Cognitive Neuroscience* 19 (9): 1498–1507. https://doi.org/10.1162/jocn.2007.19.9.1498.

4. Marcus, Daniel S., Tracy H. Wang, Jamie Parker, John G. Csernansky, John C. Morris, and Randy L. Buckner. 2007. 'Oasis Fact Sheet: Cross-Sectional Data Across the Adult Lifespan'. https://www.oasis-brains.org/files/oasis_cross-sectional_facts.pdf.

5. Rubin, Eugene H., Martha Storandt, J. Philip Miller, Dorothy A. Kinscherf, Elizabeth A. Grant, John C. Morris, and Leonard Berg. "A prospective study of cognitive function and onset of dementia in cognitively healthy elders." *Archives of neurology* 55, no. 3 (1998): 395-401.

6. Sophia Z Liu, Ghazaal Tahmasebi, Ying Sheng, Ivo D Dinov, Dennis Tsilimingras, and Xuefeng Liu. 2023. 'Sex Difference in the Associations of Socioeconomic Status, Cognitive Function and Brain Volume with Dementia in Old Adults: Findings from the OASIS Study'. *medRxiv*, January, 2023.01.05.23284240. https://doi.org/10.1101/2023.01.05.23284240.

7. Whitwell, Jennifer L., William R. Crum, Hilary C. Watt, and Nick C. Fox. "Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging." *American journal of neuroradiology* 22, no. 8 (2001): 1483-1489.